2018-12-18

Rapid detection of extra-intestinal pathogenic Escherichia coli multi-locus sequence type 127 using a specific PCR assay.

O'Hara, RW

10.1099/jmm.0.000902
Journal of Medical Microbiology
Microbiology Society

Rapid detection of Extra-intestinal Pathogenic *Escherichia coli* Multi-locus Sequence Type 127 using a specific PCR assay

*Robert W.O'Hara[1], Peter J. Jenks[2] Matthew Emery[3], & Mathew Upton[1]**

[1]University of Plymouth, Faculty of Medicine and Dentistry, Plymouth, UK.
[2]Department of Microbiology, University Hospitals Plymouth, Derriford, Plymouth UK
[3]University or Plymouth, School of Biological and Marine Sciences, Plymouth UK.

*Corresponding author – Dr Mathew Upton

E-mail: mathew.upton@plymouth.ac.uk
Tel +44(0)1752 584466
School of Biomedical Sciences, University of Plymouth, Derriford Research Facility, Research Way, Plymouth, PL6 8BU.

Short running title: Rapid detection of ST127 UPEC

Keywords: UPEC, ST127

Abstract

Members of the ST127 uropathogenic *E. coli* (UPEC) clone have a high virulence potential based on gene carriage and they are highly virulent in insect infection models. However, strains of this lineage are reported in relatively low numbers in many studies. ST127 strains are also usually widely susceptible to antibiotics and, consequently, their true prevalence may be under-recognised, as they will be eradicated during empiric therapy. A genuine concern is the possibility that members of this highly virulent lineage will acquire resistance, leading to a more serious threat. The aim of this study was to design and validate a PCR assay specific to ST127. Genomic sequences obtained from various UPEC isolates from the leading clones were used in comparative genomics to allow identification of highly discriminative sequences specific to *E. coli* ST127. The *fliC* (flagellin) and a homologue of the *upaG* (Autotransporter adhesin) gene were identified as meeting our criteria and were used to develop a multiplex PCR assay. A total of 143 *E. coli* UPEC isolates representing 99 different MLST clones from three locations (North West and South West England and Riyadh, Saudi Arabia) were used to validate the PCR assay. The multiplex PCR readily identified all 29 *E. coli* ST127 isolates, but equally importantly, produced no false positives with representatives of any of the other 98 ST's tested. We report the design and validation of a specific multiplex PCR for the rapid and reliable identification of ST127, which can be used for enhanced surveillance for this high-risk clone.

Introduction

Escherichia coli ST127 is a recently emerged clone (1) responsible for a significant proportion of extra-intestinal infections primarily of the urinary tract (1) but it has also been implicated in blood stream infections (BSI) (2, 3) and necrotizing enterocolitis in preterm infants (4). Members of the ST127 clone possess the common uropathogenic O6 serotype (5) and display an increased lethality in comparison to the more common UPEC ST lineages (ST73, ST131, ST95) with an *in vivo* model of infection (*Galleria mellonella*) (6). Additionally, ST127 strains consistently exhibit higher scores in virulence factor PCR based assays compared to representatives of some of the more frequently encountered UPEC STs (1, 6-9). Members of the ST127 clone are often reported in relatively low, but significant proportions in prevalence studies (1, 10, 11).

In general, ST127 isolates are fully susceptible to antibiotics commonly used for the empirical treatment of urinary tract infection (UTI) (7, 11, 12). They are, therefore, likely to be relatively under-represented in most published prevalence surveys, given that such studies are frequently based on UPEC isolates collected from clinical laboratories from individuals who have failed antimicrobial therapy; empirical therapy will usually result in elimination of ST127 isolates. Studies from Europe, Canada,

Saudi Arabia and Japan (1, 3, 7, 10, 13) report ST127 at low, but often significant levels. A recent study by Yamanji and colleagues (14) focused on community acquired UTI (CA-UTI) within a Californian university community. This study found ST127 to be the second most prevalent strain increasing from 11% in 1999-2000 to 16% in 2016-2017. In light of this recent evidence and reports of emerging resistance to antibiotics, including the cephalosporins, in ST127 isolates (8, 15, 16), members of this lineage are increasingly becoming a cause for clinical concern and give this strain the potential to emerge as a significant threat to human health. Ongoing surveillance of this high-risk clone is, therefore, important.

Multi-locus Sequence Typing (MLST) has been extremely beneficial with the identification of common lineages associated with UTI and BSI. However, MLST is costly and time consuming, therefore, being impractical for the rapid identification of members of the ST127 clone. One solution to these problems is the development of ST specific PCR assays, as have been designed for other STs (17-19) and have been shown to be very useful in surveillance or in the examination of large culture collections. With the use of comparative genomics and clinical isolates, for the first time, we report the design and validation of a 3-gene multiplex PCR, incorporating an extraction/PCR control and two ST127 specific targets. The assay is unambiguous in its interpretation and highly specific to *E. coli* ST127. It is easy to perform and can be used in a clinical setting to quickly monitor the prevalence and dissemination of this recently emerged, highly virulent clone.

Materials and Methods

Strains

A total of 10 ST127 and two ST73 isolates were obtained from the clinical laboratories at Derriford hospital (University Hospitals Plymouth, UK). Each isolate was recovered from non-duplicate patient urine samples that had been referred to the laboratory for standard microbiological examination for UTI. These ST73 and ST127 isolates were selected for genome sequence analysis from a wider collection of isolates collected between April 2015 and May 2015. The two ST73 isolates were included in the study as they were both isolated in the same urine specimen as one of the ST127 strains, originally identified as a recurrent monomicrobial UTI, but subsequently identified as three individual isolates with differing antimicrobial sensitivity patterns (data not shown). Isolates were identified as *E. coli* by biochemical and MALDI Biotyper analysis (Bruker Daltonics Inc.). MLST was performed on each of the 12 Derriford isolates using the Achtman scheme (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli). In addition to the 12 isolates described above, a total of 131 previously typed UPEC isolates, representing 99 different STs, were obtained from earlier studies, 52 originating from Riyadh, Saudi Arabia (10), 78

from North West England (including the ST131 reference strain EC958) (1, 20) and the reference ST127 strain 536 (21).

A total of 29 ST127 isolates were used in the validation of our PCR screening assay. One of these was the well-documented 536 reference isolate and the remainder were clinical isolates from this and our previous studies. Fourteen isolates were confirmed as ST127 by whole genome sequencing and MLST loci Sanger sequencing, and the remaining 14 ST127 isolates identified using Sanger sequencing only. Routine cultures of all *E. coli* used in this study were grown aerobically at 37°C using Lysogeny Broth (LB) or LB agar.

Genome sequencing, assembly and annotation of ST127 isolates

Draft sequenced genomes were obtained for the 12 specimens isolated from Derriford hospital (10 ST127 and 2 ST73), 2 ST127 isolates from Saudi Arabia and 2 from the North West of England. Sequencing of the *E. coli* isolates was performed on the Illumia platform by MicrobesNG (https://microbesng.uk/). The genomes were sequenced to a depth of between 48x and 94x coverage, raw sequence data assembled into contigs using SPAdes-3.9.0 (22) and the contigs ordered and aligned with reference to the UPEC ST127 536 genome (Accession ref |NC_008253 (23)) using Mauve 2.4.0 (Darling lab, University of Technology, Sydney). Sequences were concatenated using SeqHandler v0.5 (https://github.com/happykhan/seqhandler) and annotated using prokka 1.11 (24). Assembly quality was assessed using QUAST 4.0 (25) and as an additional confirmation, genomes were uploaded to the Centre for Genomic Epidemiology MLST website (https://cge.cbs.dtu.dk/services/MLST) to corroborate the original MLST result. The predicted H antigen serotype of all sequenced ST127 isolates was determined using SerotypeFinder 1.1 (26)(https://cge.cbs.dtu.dk/services/SerotypeFinder). The complete annotated chromosomes of the sequenced ST127 and ST73 isolates are available at the European Nucleotide Archive (ENA; http://www.ebi.ac.uk/ena) (Accession Number Pending).

Comparative Genomics

In addition to the UPEC sequence data obtained from the draft genomes listed above, a selection of publicly available genomes were downloaded from the NCBI (National Library of Medicine, Bethesda, Maryland, USA) database, including those for nine reference UPEC isolates and one asymptomatic bacteriuria strain (Table 1). Blast Ring Image Generator (BRIG 0.95-dev.0004) (27) (http://sourceforge.net/projects/brig) was used initially to compare all sequenced genomes to reveal regions in the ST127 genome that appeared to be absent in the genomes of UPEC form other STs. Regions of variability were examined further using the Artemis Comparison Tool (28)(http://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act). Nucleotide and protein Blast searches (https://blast.ncbi.nlm.nih.gov,

https://www.uniprot.org/blast/) were used to confirm the putative ST127 specific genomic regions that were then targeted with PCR.

Primer design and PCR protocol

The PCR primer pairs were designed using CLC Genomics Work bench 7.5.1 (https://www.qiagenbioinformatics.com). Primer targets were based on the *upaG* autotransporter and *fliC* flagellin gene regions that showed little or no homology in non-ST127 genomes using the NCBI BLAST and UniProt databases. The well-established MLST locus *gyrB* (DNA gyrase subunit B) primers (29) were incorporated into the multiplex PCR to act as an extraction/PCR control. Primer sequences, concentrations and amplicon size are listed in Table 2. DNA template for the PCR reaction was obtained via colony PCR. Briefly, a suspension of each isolate was made using material from overnight plate cultures in nuclease free water (NORMAPUR®, BDH Chemicals, VWR) to a turbidity equivalent to a 0.5 McFarland standard. This suspension was diluted 1:50 and 1.6µl used in the final PCR reaction.

Each multiplex PCR reaction was performed using 10µl of 2X Biomix™ Red reaction mix (Bioline, London, UK) in a final PCR volume of 20 µl. A primer concentration of 1pmol/µl, 1.3pmol/µl and 0.85pmol/µl was found to be most applicable for each of the *upaG*, *fliC* and *gyrB* primers, respectively, and 0.8µl of each primer was added to the PCR reaction. PCR was performed on a T100 Thermal cycler (BIO-RAD, Hertfordshire, UK) as follows: An initial denaturation at 98°C for 8 minutes followed by 36 cycles of 95°C for 30 seconds, 58°C for 1 minute and 72°C for 40 seconds, with a final extension of 72°C for 5 minutes. PCR amplification was visualised by running 5µl of the PCR product on a 1% agarose gel. Gel images were visualised under UV transillumination and the number and size of amplicon products determined. A positive ST127 isolate was identified by the presence of 3 distinct DNA bands (Fig. 2) each of the expected amplicon sizes (Table 2) and a negative reaction by the presence of the *gyrB* band alone or with just one of the 2 specific targets amplified. Amplification of the *gyrB* gene was necessary to determine a true negative result.

Validation, sensitivity and specificity of ST127 specific PCR

To determine the sensitivity and specificity of the assay, a total of 143 strains of *E. coli* (see strains section) were used to validate the multiplex PCR. The collection consisted of 29 ST127 strains and 114 UPEC isolates representing 98 different STs (Table 3).

Results and Discussion

Identification of gene regions specific to ST127

Several regions on the ST127 genome were identified as having low homology to genomes of other UPEC STs, using BRIG comparisons (Fig. 1). Each locus that was identified with ≤70% homology underwent further investigation using blastn and blastp database comparisons, culminating with the identification of two gene variants highly specific to ST127, namely the *fliC* gene and a putative *upaG* gene.

The *fliC* gene codes for the subunit protein flagellin, the major constituent of the flagellar filament. The polymorphic and antigenic properties of flagellin have been well studied since the 1930's and form an integral part of the serological classification scheme. The flagellin proteins are conserved at terminal regions while the central region is variable and often carries an H-serotype specific epitope (30). The protein is also implicated in pathogen-associated interactions, stimulating the Toll-like 5 receptor (31) which, in turn, has given rise to the prospect of the more conserved regions of the flagellin protein becoming a potent adjuvant in the design of new vaccines for UTI (32, 33). Each of the sequenced ST127 genomes carried an identical 1668bp *fliC* gene, with the single exception in isolate SA189 (from Saudi Arabia), which exhibited a C to A substitution at position 699, producing only a synonymous mutation in a valine codon. Using SerotypeFinder 1.1 the ST127 *fliC* gene was seen to have 100% identity to the predicted serotype H31 variant, in agreement with previous reports for carriage of this serotype in ST127 strains (34). The conservation within UPEC ST127 isolates, coupled with the reported variability within the *E. coli* species and previously published studies employing *fliC* as a discriminatory marker with enteropathogenic *E. coli* (35), justified selection of *fliC* as a worthy candidate for ST127 specific PCR.

The second locus identified as a putative ST127 marker, was a large 4875bp gene sequence that shared 73% identity with the *upaG* trimeric autotransporter (AT) protein found in *E. coli* CFT073 (36). The presumptive UpaG protein sequence in ST127 was found to share many structural features with *E. coli* CFT073 UpaG, the *Yersinia yadA* and *Haemophilus influenza Hia* AT genes, including specific homologies with the Hia and YadA proteins at the C-terminal region and the Left-handed Beta-roll of YadA at the hydrophobic N-terminal region (37, 38). The 73% identity with the CFT073 UpaG is not surprising as variability within genera and species for the AT family of adhesins is particularly high. The membrane anchor is the only domain that remains homologous throughout the AT and, as such, defines the family (39). The *yadA* gene was identified in *Yersinia* species and originally named P1 (40) or autoagglutination protein (41) and to date remains the best characterised AT family of adhesins (42, 43). The AT adhesins are important virulence factors for many Gram-negative pathogens and, although they are universally associated with adherence to epithelial cells and extracellular matrix (ECM) proteins (36), their functions appear extensive with reported roles in biofilm formation (44), invasion into host cells (45) and serum resistance (46).

Specificity and sensitivity of ST127 specific PCR

Ideally, one PCR target would be used to identify this particular ST, but with the size and variability within the *E. coli* pan genome, it was believed that such a precise single PCR target may be over optimistic. However, the FliC flagellin PCR primers proved to be highly sensitive and specific for the UPEC isolates producing only three false positive results (ST372, ST420 and ST1529) and a sensitivity and specificity result of 100% and 97.3%, respectively. This suggests that the H31 serotype is less common amongst other UPEC strains and, although not exclusive to ST127, remains relatively specific. The *upaG* primers alone were less specific (92.1%) with 9 non-ST127 isolates (ST14, ST80, ST141, ST537, ST540, ST550, ST785, ST807 and ST998) producing an amplicon from the *upaG* primers. However, three of these false positive results (ST14, ST550 and ST807) were easily distinguishable as negative as they produced a shortened amplicon in the range of 1000-1100bp (Fig. 3).

The two primers used in combination gave 100% sensitivity and specificity when tested against the 143 isolates, representing strains from 99 diverse UPEC associated STs. The assay was optimised with and without the inclusion of the *gyrB* extraction/PCR control and using both purified DNA (data not shown) and colony PCR. On the rare occasion a weak *upaG* and/or *fliC* amplicon was observed during the validation, the PCR was retested with both the 3 locus multiplex PCR and with the *gyrB* primers removed from the reaction. The removal of the competing *gyrB* primers from the multiplex PCR can increase the concentration of ST127 specific target amplicons, thus enhancing visualisation of the ST127 specific bands.

Although a PCR assay targeting a hypothetical protein for the detection of ST127 has recently been published by Ciesielczuk *et al* (2), their study did not provide any assay conditions or clinical validation. Their PCR used a single locus to identify ST127 and, while our own *in silico* analysis predicts that their primers should have high specificity, with no laboratory validation of performance, the utility of this assay has not been confirmed. The three false positive results we found with our primers for *fliC* indicate that even primers which appear highly specific *in silico* may perform less well when used in practice. A significant finding of our study was that no single primer pair was able to reliably identify ST127 and to achieve this, a combination of PCR targets was required. Additionally, any specific PCR assay without the presence of an extraction/PCR control will always introduce an element of doubt upon obtaining a negative result. This will reduce the practicality for use in a diagnostic setting. Although our multiplex PCR approach successfully identified all ST127s from a large collection of UPEC STs and from the very different locations in the UK and Saudi Arabia, further confirmation of performance using isolates from other geographical locations would be of value.

Concluding remarks

The vast majority of uncomplicated CA-UTI are treated empirically leading to significant over-prescription and biasing the collections of isolates investigated in many studies, which include only isolates from clinical microbiology laboratories, i.e.

isolates from cases where empirical therapy may have failed. Historically, members of the ST127 clone have been widely susceptible to first line empiric antibiotics, so will not feature in such culture collections. To greater understand the true genetic background of aetiological agents of UTI, it will be paramount that future CA-UTI studies focus on specimens collected from all patients at the point of care prior to empiric treatment. The investigations performed by Yamaji and colleagues (14) go some way to emphasise the importance of such studies. In their work, ST127 was found to be the second most common lineage in a presumably young and healthy Californian student cohort, in contrast to reports of low prevalence from other studies involving some selection bias (1, 3, 7, 10). It is understood that the presence of antimicrobial resistance in a pathogen is a prerequisite for increased prevalence, however, in the case of Extra-intestinal Pathogenic *E. coli* (ExPEC), resistance may not be the dominant driver towards increased prevalence. Recent studies show that drug resistant and drug susceptible strains have both remained equally prevalent in UTI and BSI over the last 17 and 11 years, respectively (14, 47). Furthermore, in the case of the globally disseminated ST131 clone, acquisition of specific virulence determinants predates the mutations in the *gyrA* and *parC* genes that led to the development of fluoroquinolone resistance in Clade C2 strains carrying the CTX-M-15 ESBL gene (48). This suggests that the presence or acquisition of virulence genes in ExPEC may be the necessary precursor towards the future success of a pathogen. The high virulence potential of ST127 is of clinical concern and evidence for the increase in CA-UTI warrants increased surveillance for members of this ST.

Here, we report the first validated multiplex PCR for detection of *E. coli* multi-locus Sequence Type 127. The assay is simple, yet highly discriminatory, rapid, robust, reliable and inexpensive. The multiplex PCR can also be performed directly from individual colonies removing the need for any extraction or DNA purification protocols. We suggest that such assays have an central place in surveillance for important UPEC clones. We urge laboratories to increase surveillance for ST127 isolates, on a prospective basis, to reduce the potential impact of isolates from this virulent clone that are increasingly being shown to acquire resistance.

Acknowledgements

The authors wish to thank the management, clinical and admin staff from Derriford hospital as well as the technical staff from the University of Plymouth who were extremely helpful and supportive throughout this study.

Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk), which is supported by the BBSRC (grant number BB/L024209/1).

Funding

This work was funded through a PhD studentship to MU and PJ from the University of Plymouth, School of Biomedical Sciences.

Transparency declarations

All authors have nothing to declare.

References

1. Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ, Upton M. 2012. Population structure, virulence potential and antibiotic susceptibility of uropathogenic Escherichia coli from Northwest England. J Antimicrob Chemother 67:346-56.
2. Ciesielczuk H, Jenkins C, Chattaway M, Doumith M, Hope R, Woodford N, Wareham DW. 2016. Trends in ExPEC serogroups in the UK and their significance. Eur J Clin Microbiol Infect Dis doi:10.1007/s10096-016-2707-8.
3. Adams-Sapper S, Diep BA, Perdreau-Remington F, Riley LW. 2013. Clonal composition and community clustering of drug-susceptible and -resistant Escherichia coli isolates from bloodstream infections. Antimicrob Agents Chemother 57:490-7.
4. Ward DV, Scholz M, Zolfo M, Taft DH, Schibler KR, Tett A, Segata N, Morrow AL. 2016. Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic E. coli in Necrotizing Enterocolitis and Mortality in Preterm Infants. Cell Rep 14:2912-24.
5. Johnson JR. 1991. Virulence factors in Escherichia coli urinary tract infection. Clin Microbiol Rev 4:80-128.
6. Alghoribi MF, Gibreel TM, Dodgson AR, Beatson SA, Upton M. 2014. Galleria mellonella infection model demonstrates high lethality of ST69 and ST127 uropathogenic E. coli. PLoS One 9:e101547.
7. Croxall G, Hale J, Weston V, Manning G, Cheetham P, Achtman M, McNally A. 2011. Molecular epidemiology of extraintestinal pathogenic Escherichia coli isolates from a regional cohort of elderly patients highlights the prevalence of ST131 strains with increased antimicrobial resistance in both community and hospital care settings. J Antimicrob Chemother 66:2501-8.
8. Beyrouthy R, Robin F, Cougnoux A, Dalmasso G, Darfeuille-Michaud A, Mallat H, Dabboussi F, Hamze M, Bonnet R. 2013. Chromosome-mediated OXA-48 carbapenemase in highly virulent Escherichia coli. J Antimicrob Chemother 68:1558-61.
9. Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B, Butler-Wu SM, Lee C, Cookson BT, Shendure J. 2015. Large-scale genomic sequencing of extraintestinal pathogenic Escherichia coli strains. Genome Res 25:119-28.
10. Alghoribi MF, Gibreel TM, Farnham G, Al Johani SM, Balkhy HH, Upton M. 2015. Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic Escherichia coli clones in Riyadh, Saudi Arabia. J Antimicrob Chemother 70:2757-62.
11. Banerjee R, Johnston B, Lohse C, Chattopadhyay S, Tchesnokova V, Sokurenko EV, Johnson JR. 2013. The clonal distribution and diversity of extraintestinal Escherichia coli isolates vary according to patient characteristics. Antimicrob Agents Chemother 57:5912-7.
12. McNulty C. 2017. Public Health England, Management and treatment of common infections: guidance for consultation and adaptation.

13. Umene YD, Wong LK, Satoh T, Yamane K, Matsui M, Riley LW, Arakawa Y, Suzuki S. 2015. Molecular epidemiological characterization of uropathogenic escherichia coli from an outpatient urology clinic in rural Japan. J Clin Microbiol 53:681-3.
14. Yamaji R, Rubin J, Thys E, Friedman CR, Riley LW. 2018. Persistent pandemic lineages of uropathogenic Escherichia coli in a college community-1999-2017. J Clin Microbiol doi:10.1128/jcm.01834-17.
15. Ferjani S, Saidani M, Hamzaoui Z, Alonso CA, Torres C, Maamar E, Slim AF, Boutiba BB. 2017. Community fecal carriage of broad-spectrum cephalosporin-resistant Escherichia coli in Tunisian children. Diagn Microbiol Infect Dis 87:188-192.
16. Gomi R, Matsuda T, Matsumura Y, Yamamoto M, Tanaka M, Ichiyama S, Yoneda M. 2017. Whole-Genome Analysis of Antimicrobial-Resistant and Extraintestinal Pathogenic Escherichia coli in River Water. Appl Environ Microbiol 83.
17. Li D, Liu B, Chen M, Guo D, Guo X, Liu F, Feng L, Wang L. 2010. A multiplex PCR method to detect 14 Escherichia coli serogroups associated with urinary tract infections. J Microbiol Methods 82:71-7.
18. Doumith M, Day M, Ciesielczuk H, Hope R, Underwood A, Reynolds R, Wain J, Livermore DM, Woodford N. 2015. Rapid identification of major Escherichia coli sequence types causing urinary tract and bloodstream infections. J Clin Microbiol 53:160-6.
19. Clermont O, Dhanji H, Upton M, Gibreel T, Fox A, Boyd D, Mulvey MR, Nordmann P, Ruppe E, Sarthou JL, Frank T, Vimont S, Arlet G, Branger C, Woodford N, Denamur E. 2009. Rapid detection of the O25b-ST131 clone of Escherichia coli encompassing the CTX-M-15-producing strains. J Antimicrob Chemother 64:274-7.
20. Lau SH, Reddy S, Cheesbrough J, Bolton FJ, Willshaw G, Cheasty T, Fox AJ, Upton M. 2008. Major uropathogenic Escherichia coli strain isolated in the northwest of England identified by multilocus sequence typing. J Clin Microbiol 46:1076-80.
21. Hacker J, Hughes C. 1985. Genetics of Escherichia coli hemolysin. Curr Top Microbiol Immunol 118:139-62.
22. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 20:714-37.
23. Hochhut B, Wilde C, Balling G, Middendorf B, Dobrindt U, Brzuszkiewicz E, Gottschalk G, Carniel E, Hacker J. 2006. Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic Escherichia coli strain 536. Mol Microbiol 61:584-95.
24. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.
25. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-5.
26. Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. J Clin Microbiol 53:2410-26.
27. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:402.
28. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. 2005. ACT: the Artemis Comparison Tool. Bioinformatics 21:3422-3.
29. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. 2006. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol 60:1136-51.
30. Wang L, Rothemund D, Curd H, Reeves PR. 2003. Species-wide variation in the Escherichia coli flagellin (H-antigen) gene. J Bacteriol 185:2936-43.
31. Smith KD, Andersen-Nissen E, Hayashi F, Strobe K, Bergman MA, Barrett SL, Cookson BT, Aderem A. 2003. Toll-like receptor 5 recognizes a conserved site on

flagellin required for protofilament formation and bacterial motility. Nat Immunol 4:1247-53.
32. Savar NS, Jahanian-Najafabadi A, Mahdavi M, Shokrgozar MA, Jafari A, Bouzari S. 2014. In silico and in vivo studies of truncated forms of flagellin (FliC) of enteroaggregative Escherichia coli fused to FimH from uropathogenic Escherichia coli as a vaccine candidate against urinary tract infections. J Biotechnol 175:31-7.
33. Bennett KM, Gorham RD, Jr., Gusti V, Trinh L, Morikis D, Lo DD. 2015. Hybrid flagellin as a T cell independent vaccine scaffold. BMC Biotechnol 15:71.
34. Johnson JR, Johnston B, Clabots CR, Kuskowski MA, Roberts E, DebRoy C. 2008. Virulence genotypes and phylogenetic background of Escherichia coli serogroup O6 isolates from humans, dogs, and cats. J Clin Microbiol 46:417-22.
35. Beutin L, Delannoy S, Fach P. 2015. Sequence Variations in the Flagellar Antigen Genes fliCH25 and fliCH28 of Escherichia coli and Their Use in Identification and Characterization of Enterohemorrhagic E. coli (EHEC) O145:H25 and O145:H28. PLoS One 10:e0126749.
36. Valle J, Mabbett AN, Ulett GC, Toledo-Arana A, Wecker K, Totsika M, Schembri MA, Ghigo JM, Beloin C. 2008. UpaG, a new member of the trimeric autotransporter family of adhesins in uropathogenic Escherichia coli. J Bacteriol 190:4147-61.
37. Szczesny P, Lupas A. 2008. Domain annotation of trimeric autotransporter adhesins--daTAA. Bioinformatics 24:1251-6.
38. Nummelin H, Merckel MC, Leo JC, Lankinen H, Skurnik M, Goldman A. 2004. The Yersinia adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. Embo j 23:701-11.
39. Linke D, Riess T, Autenrieth IB, Lupas A, Kempf VA. 2006. Trimeric autotransporter adhesins: variable structure, common function. Trends Microbiol 14:264-70.
40. Bolin I, Norlander L, Wolf-Watz H. 1982. Temperature-inducible outer membrane protein of Yersinia pseudotuberculosis and Yersinia enterocolitica is associated with the virulence plasmid. Infect Immun 37:506-12.
41. Bolin I, Wolf-Watz H. 1984. Molecular cloning of the temperature-inducible outer membrane protein 1 of Yersinia pseudotuberculosis. Infect Immun 43:72-8.
42. El Tahir Y, Skurnik M. 2001. YadA, the multifaceted Yersinia adhesin. Int J Med Microbiol 291:209-18.
43. Muhlenkamp M, Oberhettinger P, Leo JC, Linke D, Schutz MS. 2015. Yersinia adhesin A (YadA)--beauty & beast. Int J Med Microbiol 305:252-8.
44. Allsopp LP, Beloin C, Moriel DG, Totsika M, Ghigo JM, Schembri MA. 2012. Functional heterogeneity of the UpaH autotransporter protein from uropathogenic Escherichia coli. J Bacteriol 194:5769-82.
45. Eitel J, Dersch P. 2002. The YadA protein of Yersinia pseudotuberculosis mediates high-efficiency uptake into human cells under environmental conditions in which invasin is repressed. Infect Immun 70:4880-91.
46. Biedzka-Sarek M, Venho R, Skurnik M. 2005. Role of YadA, Ail, and Lipopolysaccharide in Serum Resistance of Yersinia enterocolitica Serotype O:3. Infect Immun 73:2232-44.
47. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, Peacock SJ, Parkhill J. 2017. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res doi:10.1101/gr.216606.116.
48. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook M, Schembri MA, Beatson SA. 2016. Sequential Acquisition of Virulence and Fluoroquinolone Resistance Has Shaped the Evolution of Escherichia coli ST131. MBio 7:e00347-16.

Table headings and figure legends

Table 1. List of *E. coli* strains and their designated sequence types recovered from the NCBI database and used for comparative genomics.

Table 2. Details of primers used in multiplex PCR to detect members of the ST127 clone

Table 3. List of all UPEC isolates and their designated sequence types that were used for validation of the ST127 specific PCR assay

Figure 1. Comparison of the genomes of 25 UPEC and one asymptomatic bacteriuria *E. coli* isolate. The inner circle represents the reference sequence EC536 with the inner red rings representing genomes of ST127, 10 originating from Derriford hospital, 2 from Saudi Arabia and 2 from North West England. The green circles represent non-ST127 isolates, 2 ST73 isolates from Derriford hospital and 9 genomes download from the NCBI database. Blank gaps in the rings represent ≤70% homology and shaded areas represent ≤90% homology. The red arrows indicate the position of *fliC* and u*paG* genes. The image was prepared using Blast Ring Image Generator.

Figure 2. PCR detection for members of the ST127 clone. The *gyrB* band (911bp) corresponds to the Extraction/PCR control amplified in all isolates. The *upaG* and *fliC* bands both amplified in the same reaction correspond only to isolates from the ST127 clone strains. Lanes 1-6 consisted of ST127 isolates, lanes 7-9 non-ST127 isolates. Lane 1, NW41; lane 2, D3; lane 3, D4; lane 4, NW72; lane 5, NW112; lane 6, NW154; lane 7, NW153 (ST1529 *fliC*+), lane 8, SA027 (ST537 *upaG*+); Lane 9, EC958 (ST131) and M, 1kb molecular weight marker (Bioline, London, UK).

Figure 3. Routine testing of *upaG-gyrB-fliC* ST127 specific PCR. Lanes 3, 9, 10, 13-16 and 19 were ST127 isolates, positive for 3 gene loci. Lanes 1-2, 4-8, 11-12, 17-18 and 20-24 were non-ST127 isolates. Lanes 20 and 23 (ST80 and ST550, respectively) are examples of a weak *upaG* positive (lane 20) and a *upaG* positive (lane 23) of reduced amplicon length causing the *upaG* amplicon to merge with the *gyrB* amplicon. M, 1kb molecular weight marker.

497 Table 1.

UPEC Isolate	Sequence Type	NCBI Bioproject Accession No. / Reference sequence
ABU 83972	ST73	PRJNA38725
CFT073	ST73	PRJNA313
NA114	ST131	PRJNA66975
EC958	ST131	NZ_HG941718.1
UMN026	ST597	PRJNA33415
IAI39	ST62	PRJNA33411
EC536	ST127	NC_008253
UTI89	ST95	PRJNA16259
VR50	ST10	PRJEA61445
clone D i2	ST73	PRJNA52021

498

499

500 Table 2.

Gene	Primer direction	Primer Sequence (5'-3')	Final primer concentration (µM)	Product Length (bp)	Reference
upaG	Forward	GATAGGCAAGGACGCAAGA	0.04	1218	This study
	Reverse	GGTCGCAATATCCGTAGT	0.04		This study
fliC	Forward	CATTAATACCAACAGCCTC	0.052	538	This study
	Reverse	TATTAGCCACAGCCCCTT	0.052		This study
gyrB	Forward	TCGGCGACACGGATGACGGC	0.034	911	(29)
	Reverse	ATCAGGCCTTCACGCGCATC	0.034		(29)

501

502

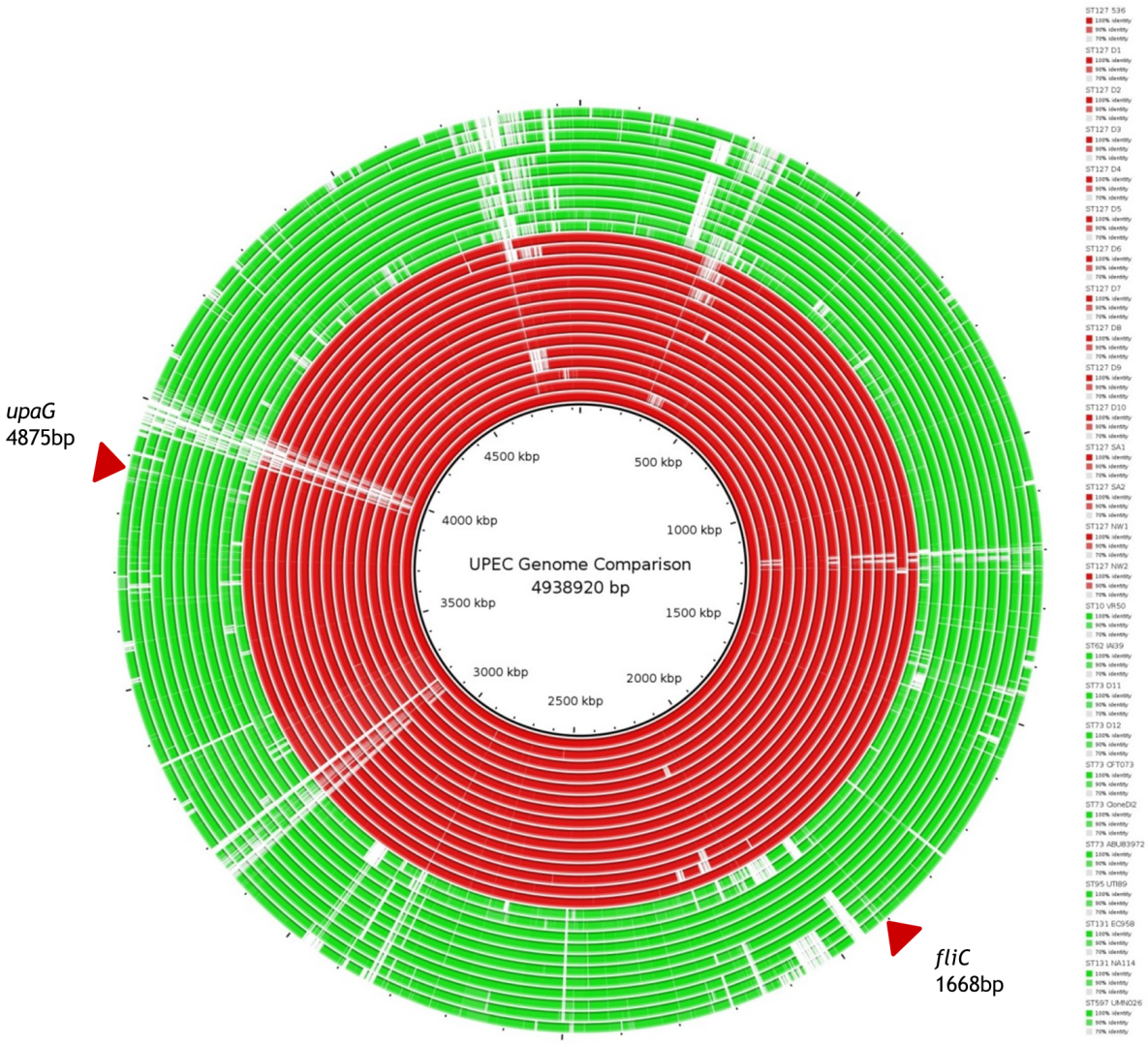
503 Table 3.

E. coli Isolate*	MLST	*E. coli* Isolate	MLST	*E. coli* Isolate	MLST	*E. coli* Isolate	MLST	*E. coli* Isolate	MLST
SA014	10	NW45	104	NW224	127	SA067	624	SA060	1380
SA036	10	NW155	117	EC958	131	SA119	662	SA061	1431
SA097	10	D9	127	NW002	131	NW178	681	NW153	1529
SA098	10	D39	127	NW014	131	NW130	779	NW169	1532
SA135	10	D124A	127	NW59	136	NW140	780	NW174	1533
SA140	10	D124B	127	NW82	141	NW141	781	NW192	1534
NW006	10	D263A	127	SA053	153	NW57	782	NW193	1535
NW212	14	D264	127	SA005	155	NW016	783	NW203	1536
SA004	23	D298	127	SA013	162	NW007	784	NW220	1537
SA023	38	D316	127	NW39	224	NW63	785	NW221	1538
SA034	38	D354	127	SA082	315	NW74	786	NW225	1540
NW65	48	D468	127	SA035	347	NW34	787	NW226	1541
SA072	52	EC536	127	SA012	367	NW36	804	NW235	1542
SA217	57	NW18	127	SA010	371	NW40	805	NW236	1543
NW38	58	NW41	127	NW27	372	NW41	806	NW237	1544
NW79	59	SA009	127	NW80	393	NW42	807	NW244	1545
NW56	62	SA028	127	NW175	399	NW43	808	NW245	1546
NW019	69	SA033	127	NW53	405	NW47	809	NW245	1547
D263B	73	SA039	127	NW98	410	SA042	998	SA062	1611
D263C	73	SA126	127	NW165	420	NW011	999	SA137	2020
NW009	73	SA151	127	NW87	448	NW49	1000	SA109	2659
NW012	73	SA153	127	SA161	449	NW61	1001	SA198	3076
NW013	73	SA168	127	SA026	450	NW71	1002	SA030	3556
NW015	73	SA174	127	NW179	493	NW78	1003		
NW001	80	SA189	127	SA132	501	NW86	1004		
NW017	88	SA191	127	SA027	537	NW88	1005		
NW003	95	SA218	127	SA063	540	SA167	1196		
NW33	95	NW72	127	SA099	543	SA123	1266		
SA056	101	NW112	127	NW004	550	NW171	1303		
SA081	101	NW154	127	SA116	617	SA142	1312		

504 *The prefix determines the location where the isolate originated from, D; Derriford
505 hospital, SA; Saudi Arabia and NW; North West England, EC; reference strains.

506

507



508

509 Figure 1.

510

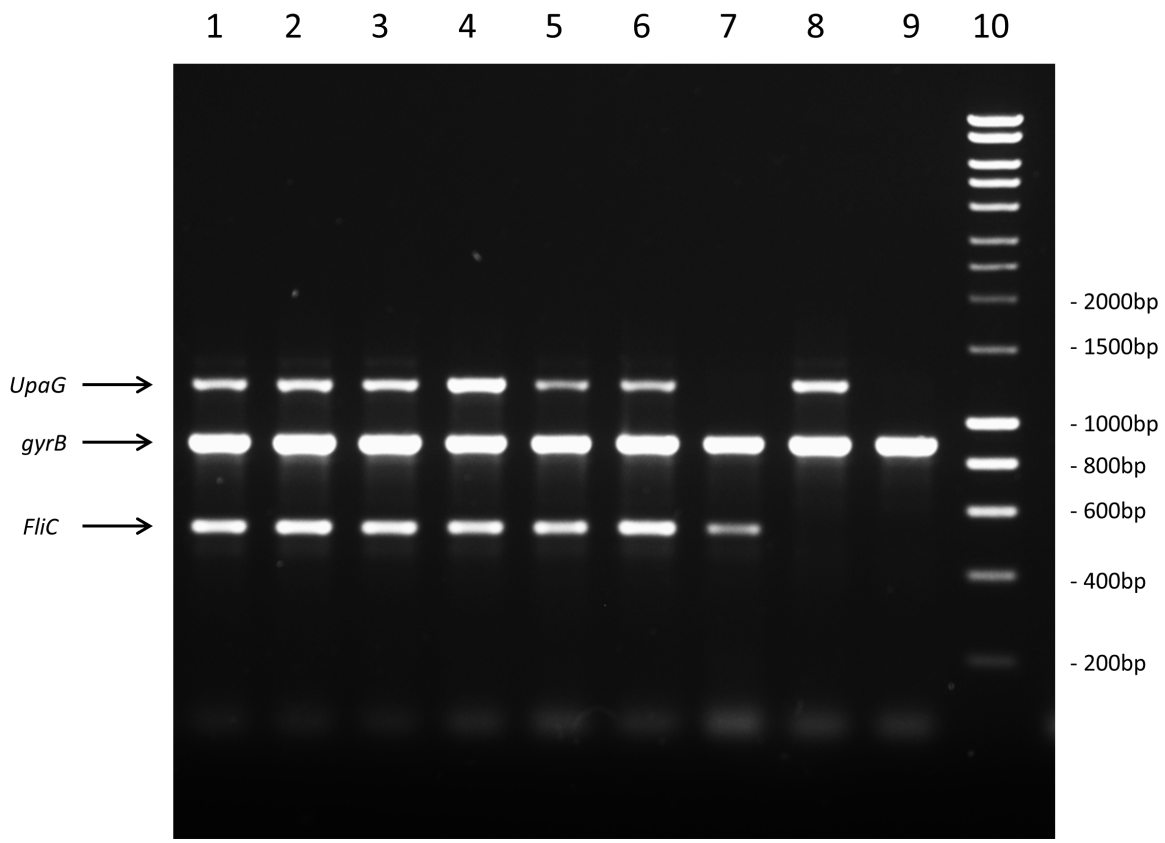


Figure 2.

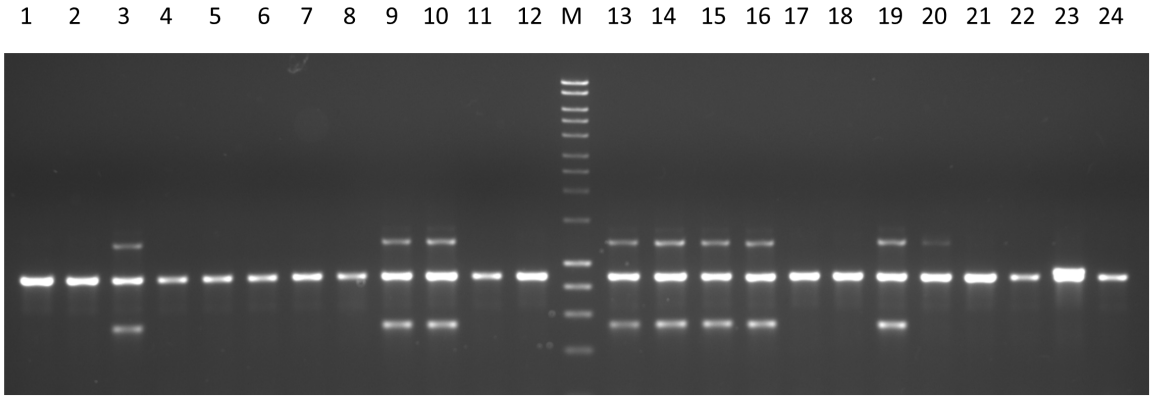


Figure 3.